Multiparent populations & R/qtl2

Karl Broman

Biostatistics and Medical Informatics
University of Wisconsin – Madison

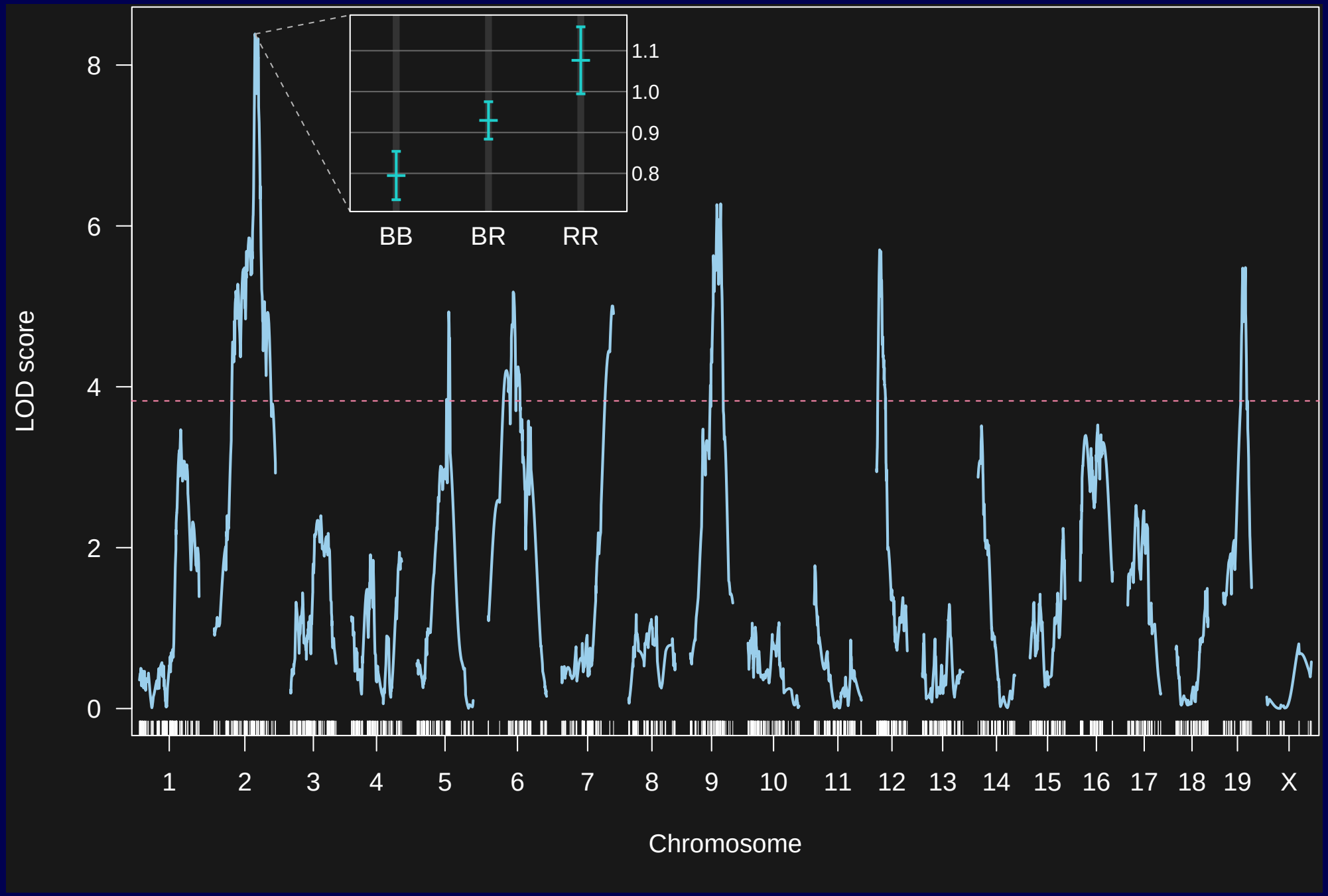
kbroman.org/qtl2

kbroman.org

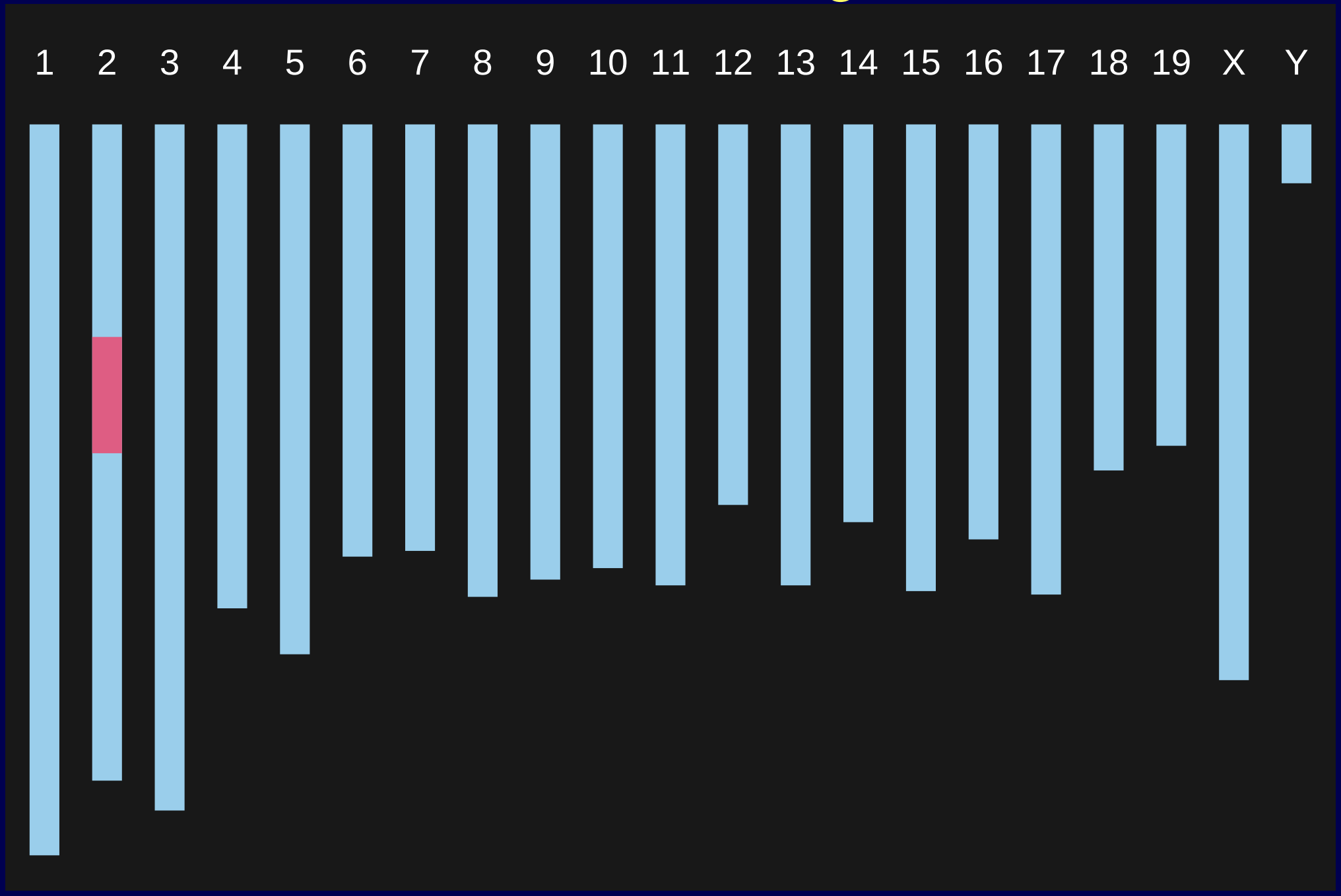
github.com/kbroman

@kwbroman

QTL mapping


LOD score
0
2
4
6
8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Chromosome
1.1
1.0
0.9
0.8
BB
BR
RR

Congenic line / NIL

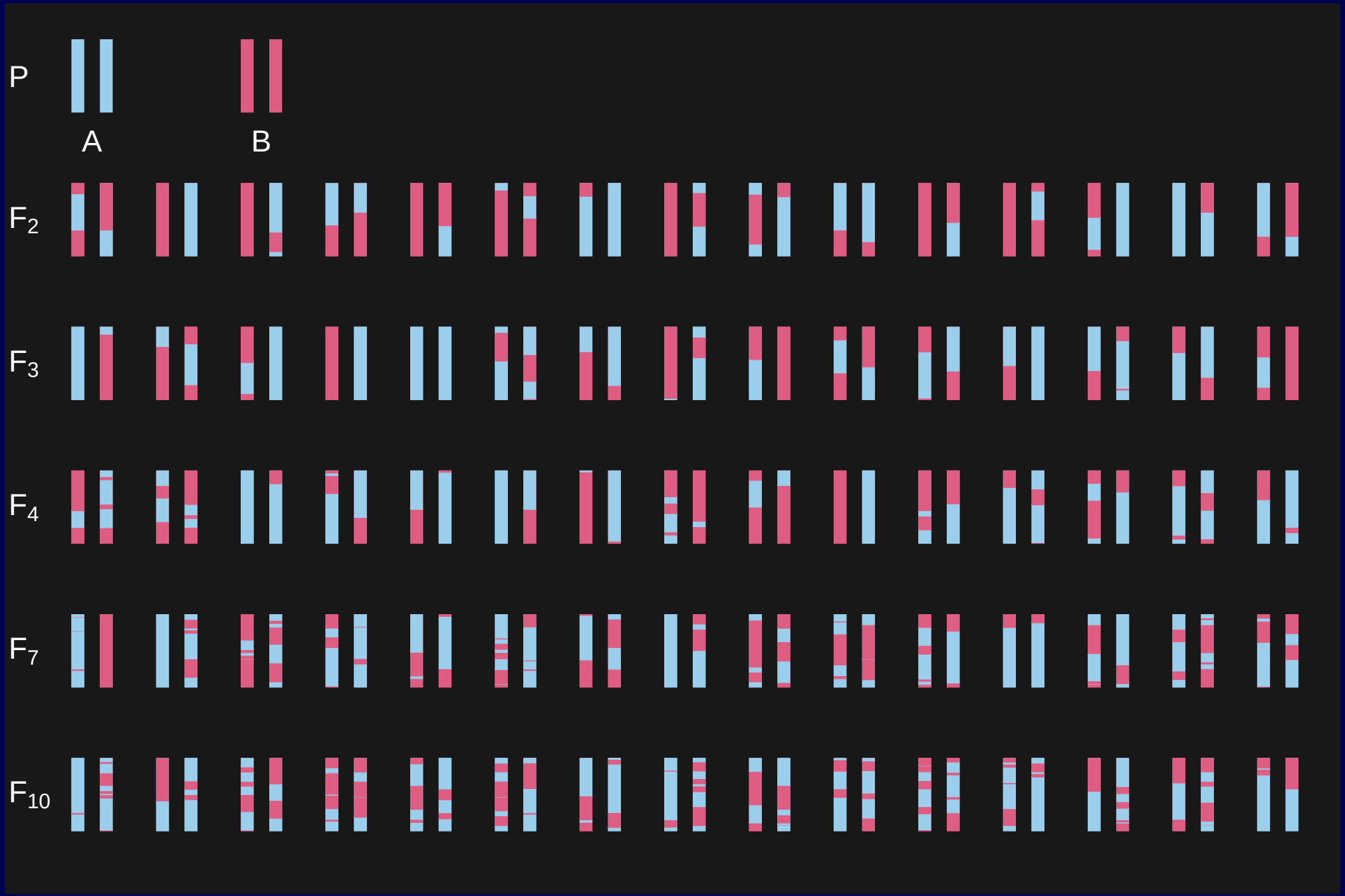


Improving precision

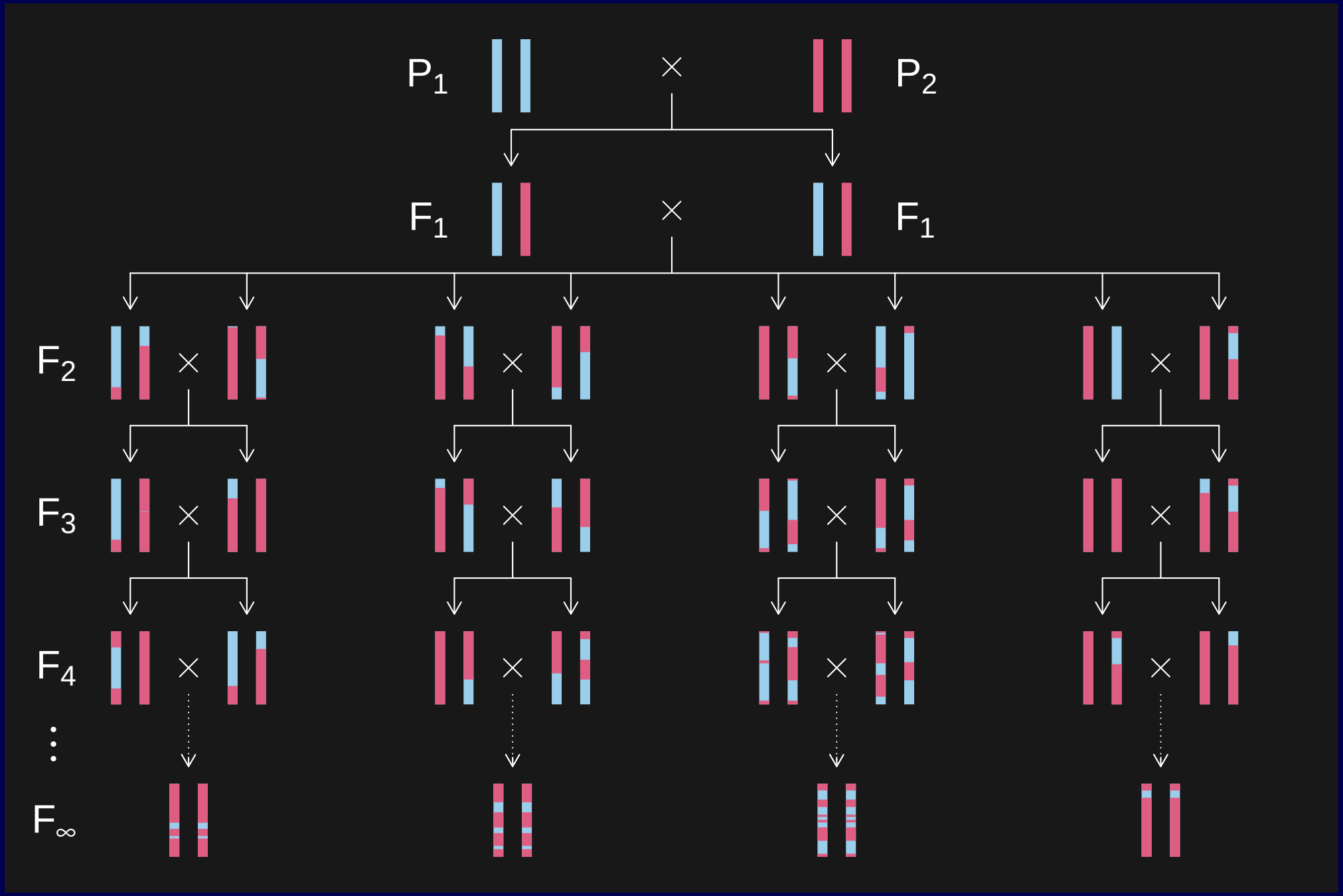
- more recombinations
- more individuals
- more precise phenotypes
- lower-level phenotypes

  transcripts, proteins, metabolites

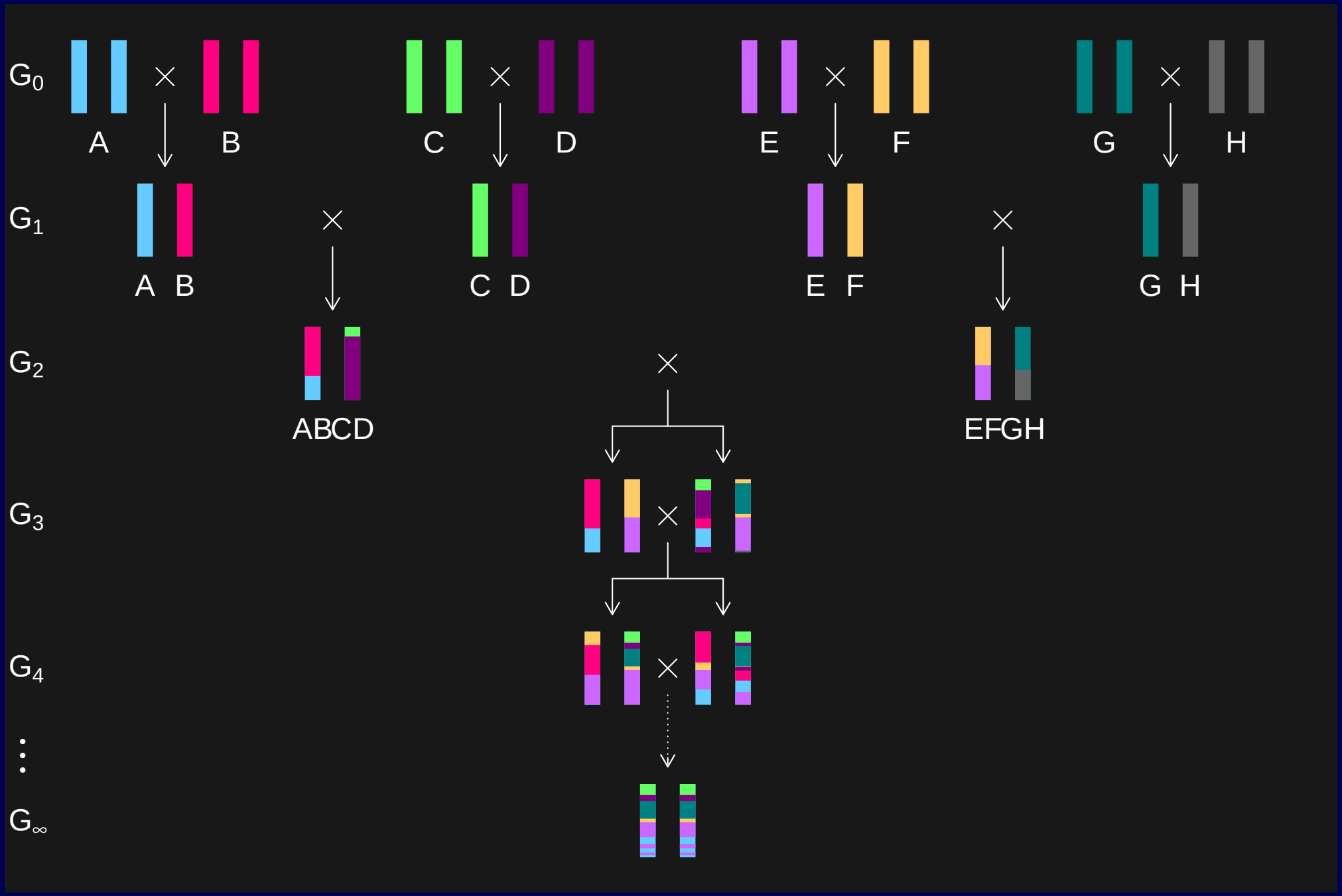
Advanced intercross lines



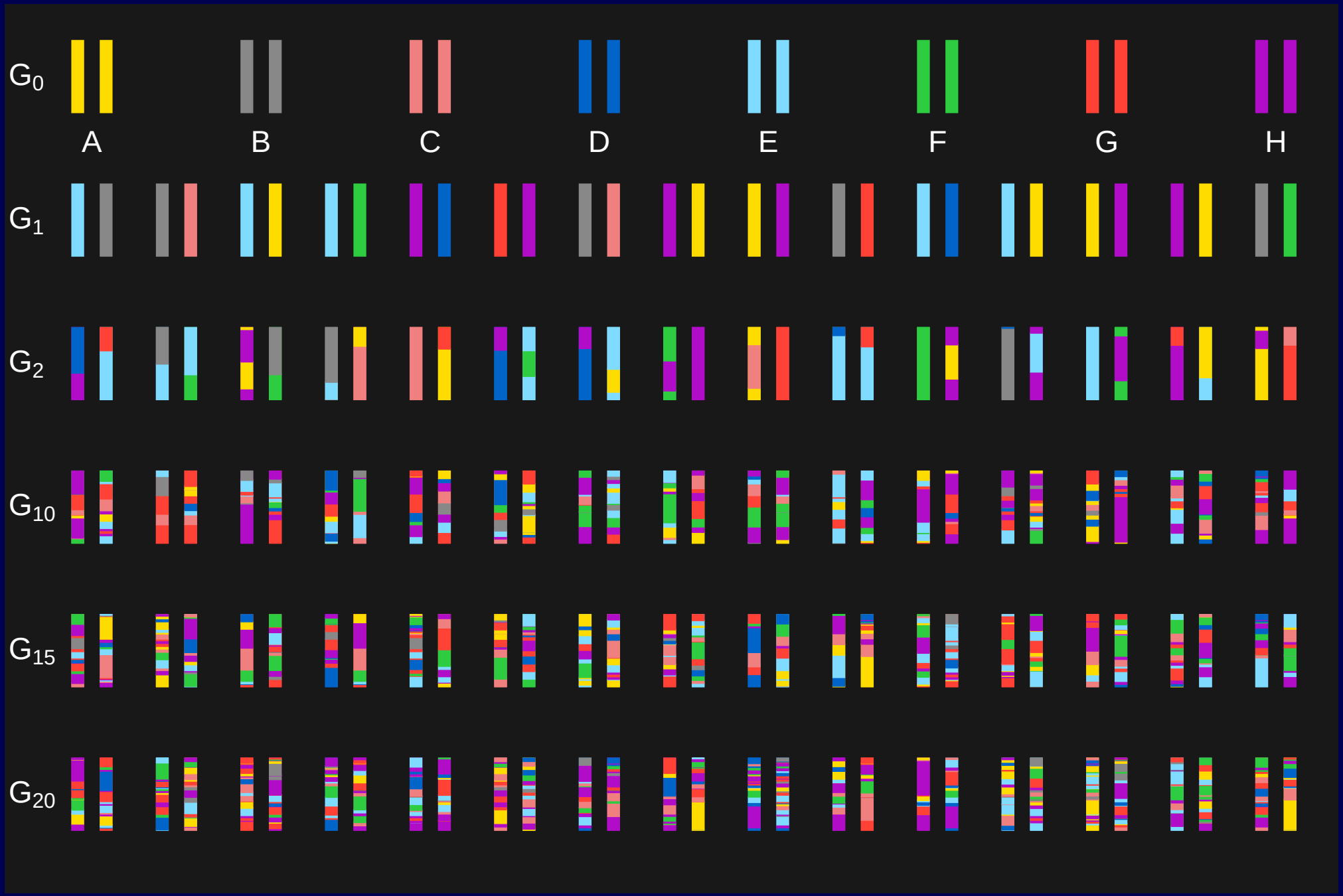
Recombinant inbred lines



Collaborative Cross/MAGIC



Heterogeneous stock



Why R/qtl2?

- High-dimensional data
  - genotypes and phenotypes
- More diverse crosses
  - especially multi-parent populations
- Linear mixed models
  - especially in DO/HS/AIL

R/qtl → R/qtl2

- See `kbroman.org/qtl2/assets/vignettes/rqtl_diff.html`
- New data file formats
- New data structures
- New function names

  `read.cross()` → `read_cross2()`

  `calc.genoprob()` → `calc_genoprob()`

  `scanone()` → `scan1()`
- Different treatment of intermediate calculations
- Use of individual IDs for aligning data
- Order of args when subsetting cross objects

  `cross[chr,ind]` → `cross2[ind,chr]`

→ R

- `convert2cross2()`
- `summary()`, `n_ind()`, `n_mar()`, ...
- `insert_pseudomarkers()`
- `calc_genoprob()`
- `scan1()`
- `find_peaks()`

Linear mixed models

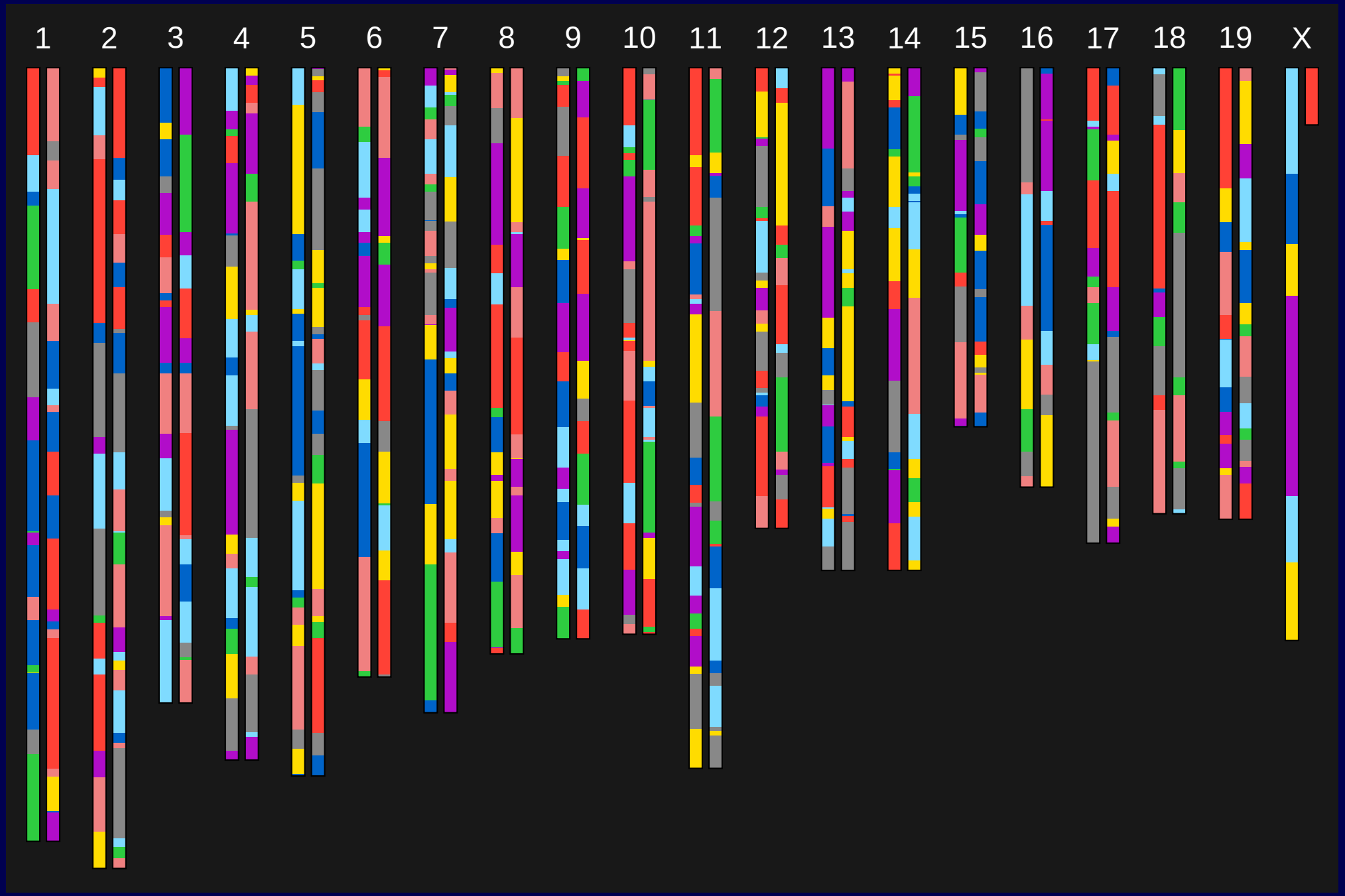
$$
\begin{aligned}
y_i &= \mu + \sum\nolimits_k \beta_k q_{ik} + \epsilon_i \qquad \epsilon_i \sim \mathsf{N}(0, \sigma_e^2) \\
&= \mu + \qquad \eta_i \qquad + \epsilon_i \qquad \eta_i \sim \mathsf{N}(0, \sigma_p^2)
\end{aligned}
$$

$$
\mathsf{COV}(\eta_i, \eta_j) = \sigma_p^2 \, (2k_{ij})
$$

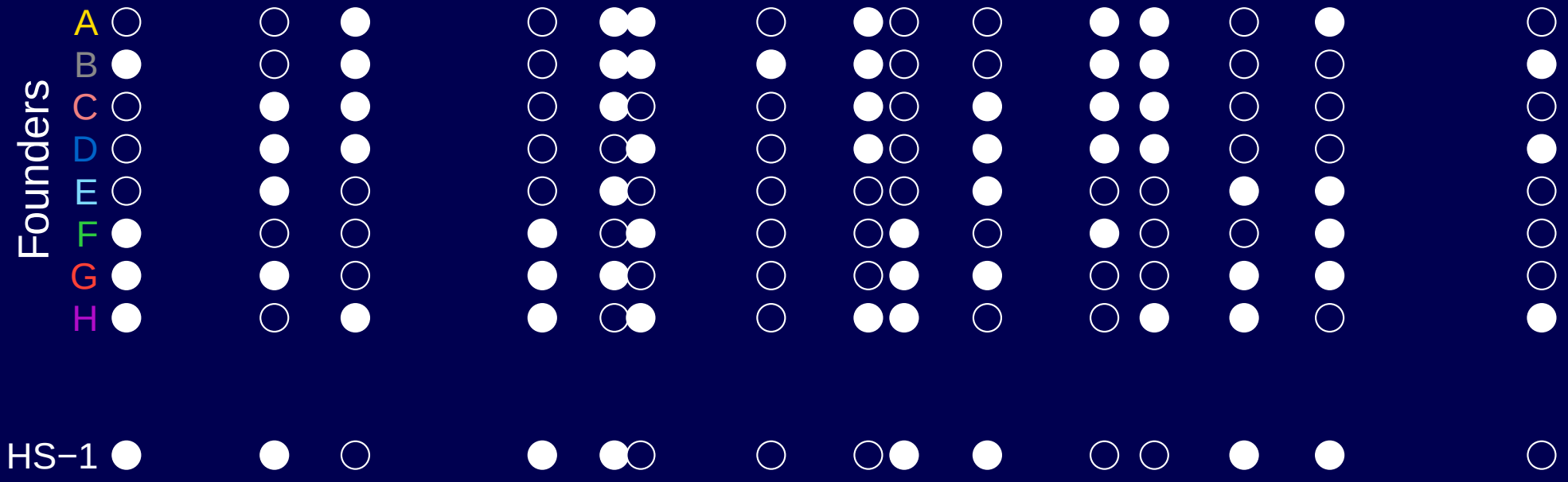
→ R

- `calc_kinship()`
- `scan1()`

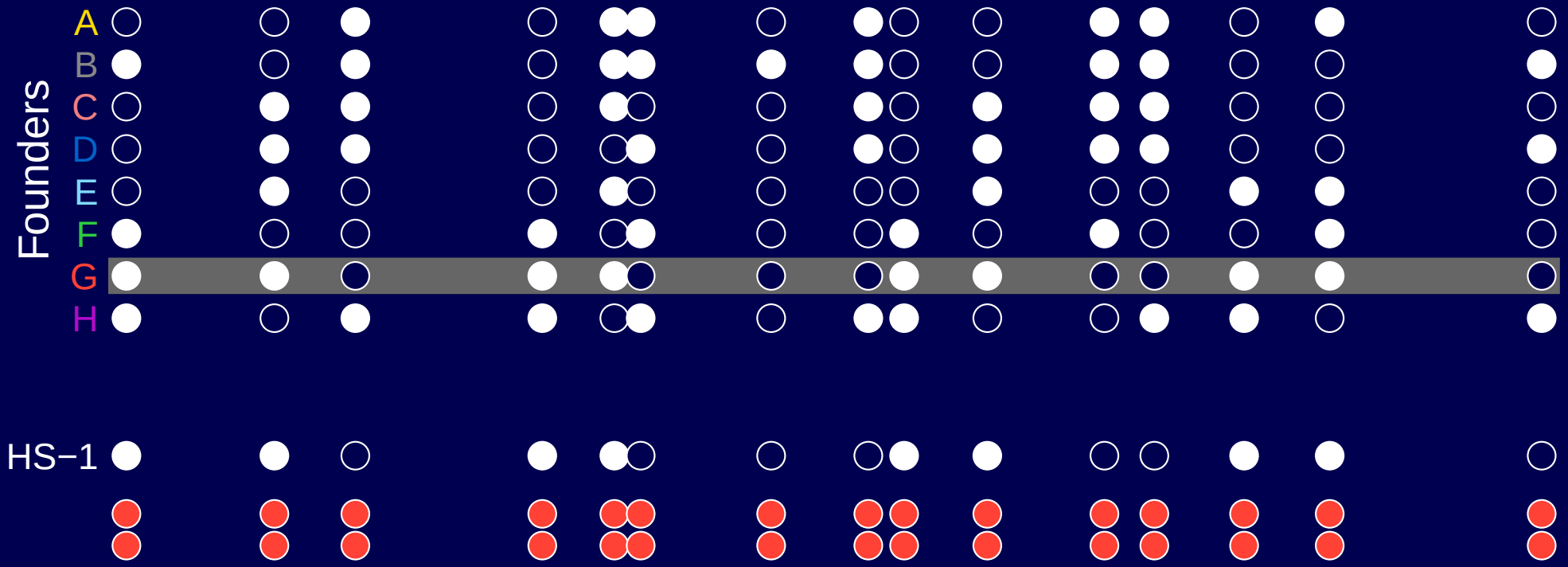
HS genome



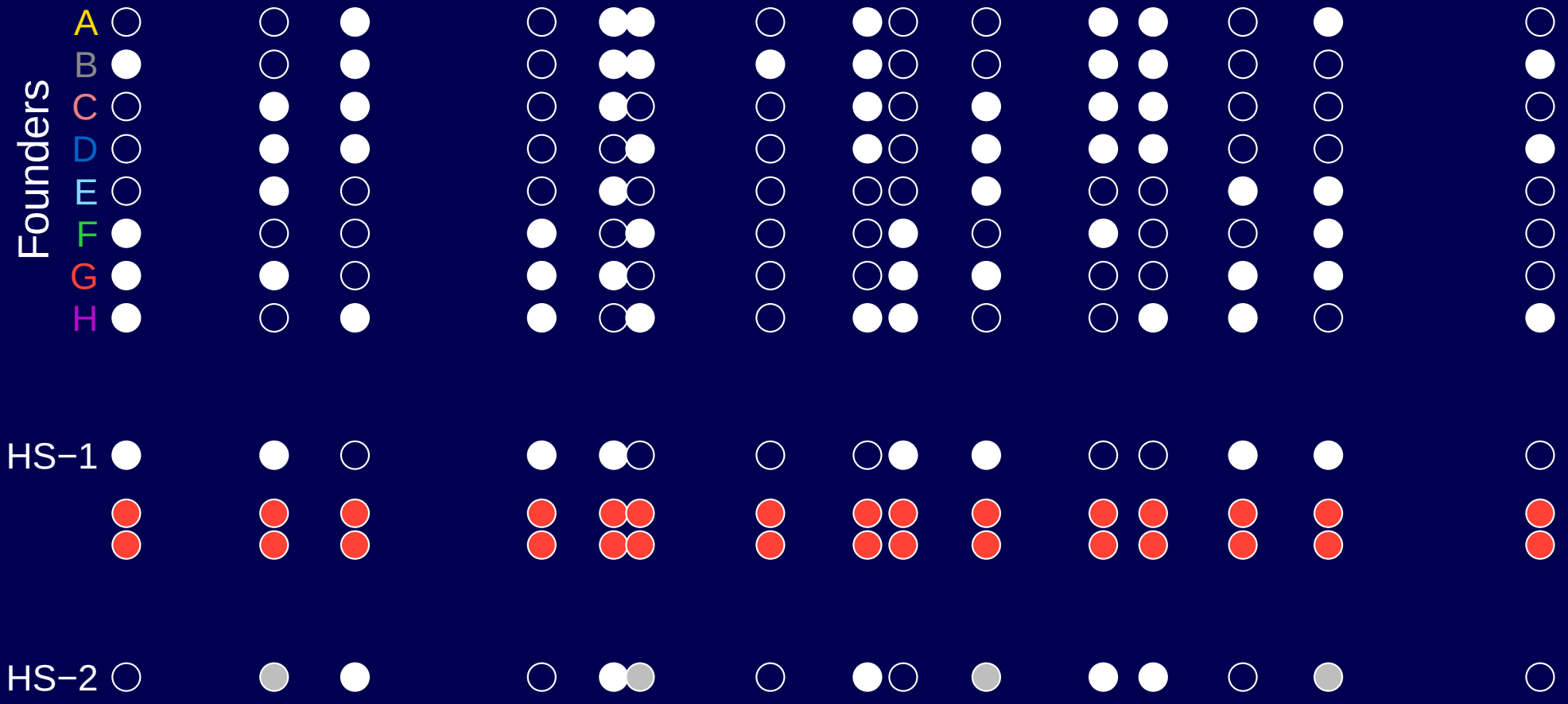
HS genotype reconstruction



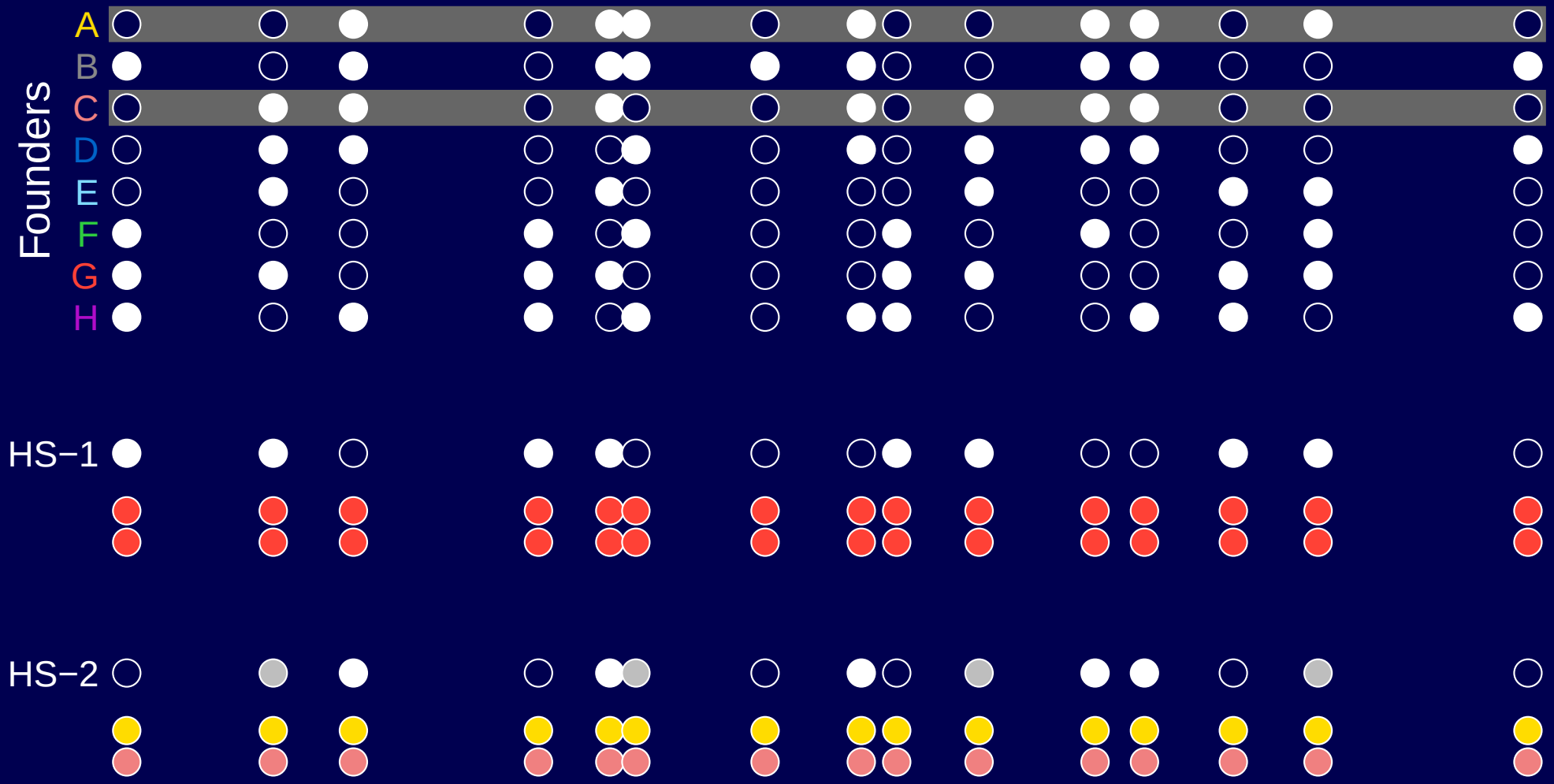
HS genotype reconstruction



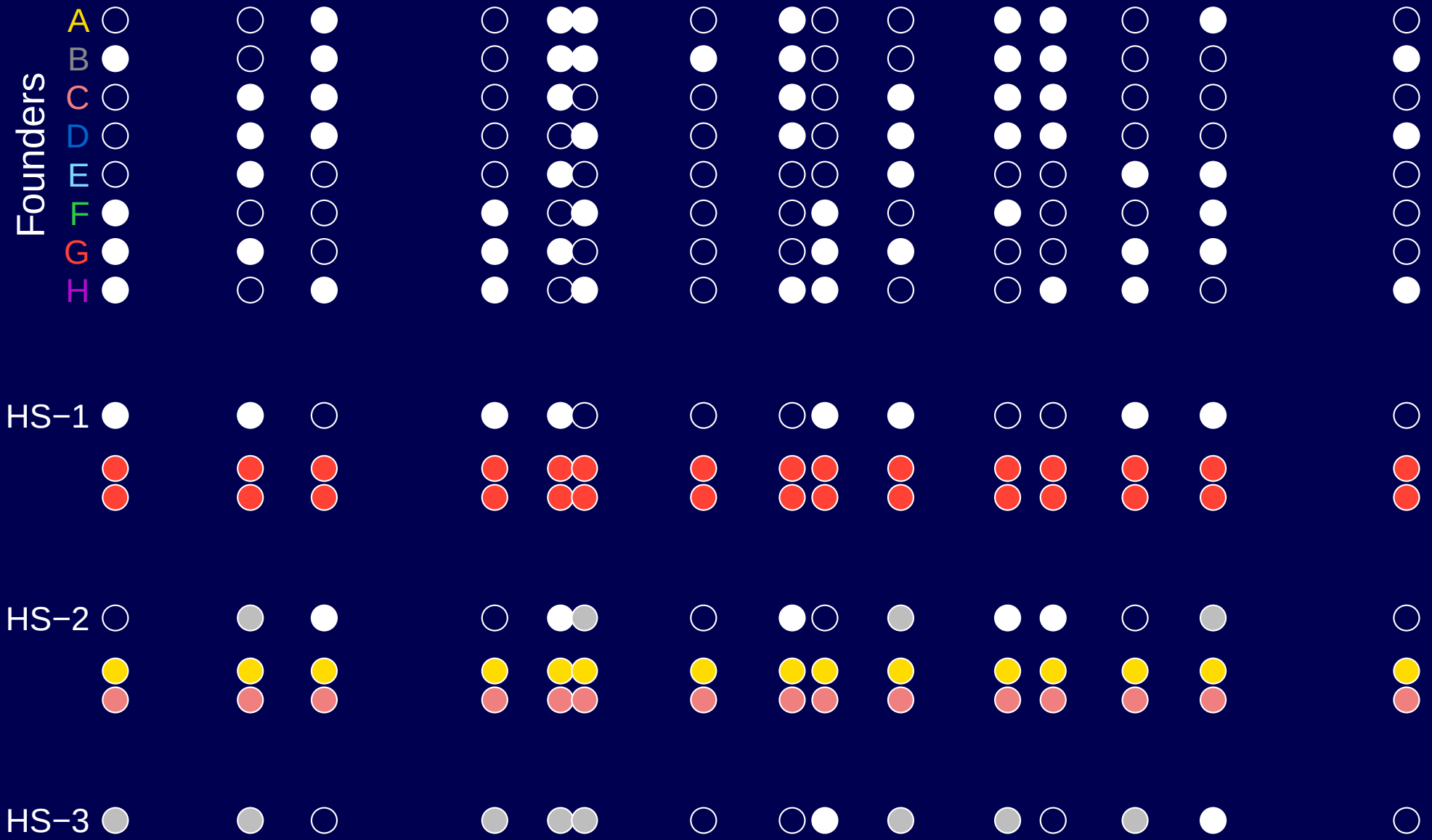
HS genotype reconstruction



HS genotype reconstruction



HS genotype reconstruction



HS genotype reconstruction